44

Validating Rainfall Parameterizations in Climate Models using Predictive Empirical Analysis

R. Saravanan (with J. Wang, R. K. W. Wong, M. Jun, and C. Schumacher)

T3: TEXAS A&M TRIADS FOR TRANSFORMATION
A President's Excellence Fund Initiative

Motivation

- Climate models exhibit significant biases in their rainfall simulation (Fig. 1)
- Moist convection parameterizations that are physically motivated have failed to fully address this problem
- Why not try a purely data-driven, i.e., statistical or machine-learning (ML) approach?

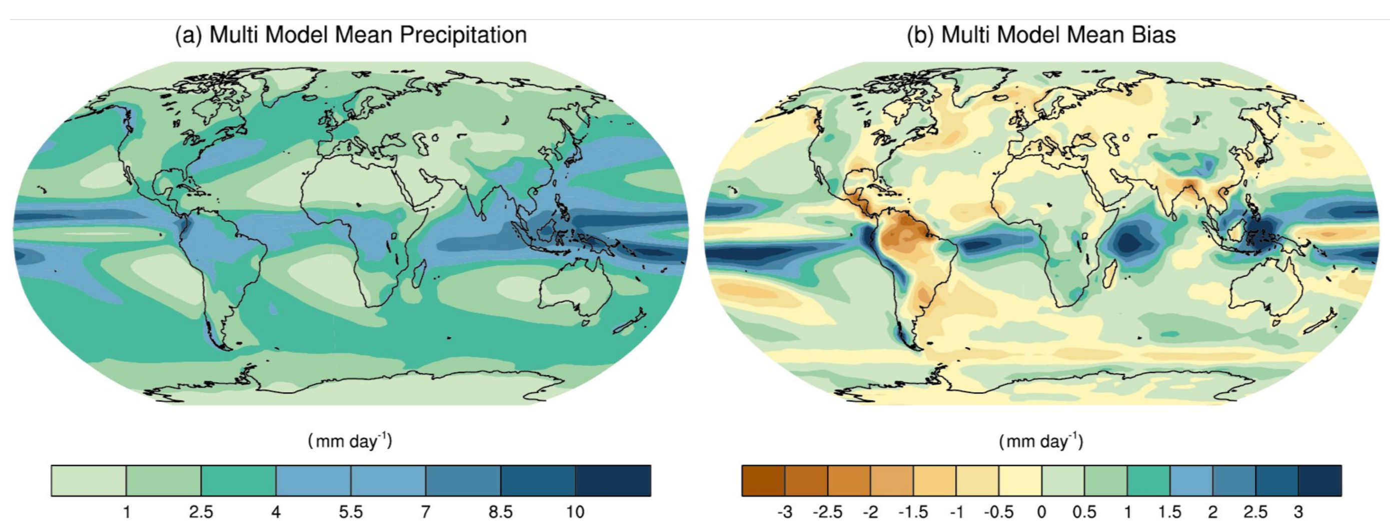
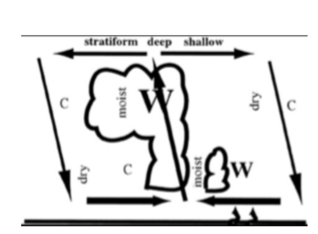


Figure 1. IPCC AR5 estimate of multi model mean precipitation and bias

Approach

- Use available data to construct a predictive model of rainfall using different statistical/ML approaches
  - GPM satellite radar rainfall data
  - MERRA 2 reanalysis for atmospheric state
  - Interpolate all data to 0.5°, 3-hourly grid
- Divide rain into 3 types
  - Stratiform (STR)
  - Deep convective (DC)
  - Shallow convective (SC)



- Initial focus on East Pacific (EP) and West Pacific (WP) regions (Fig. 2)
  - Use one year of data (2017) to train model
  - Another year of data (2018) to validate model

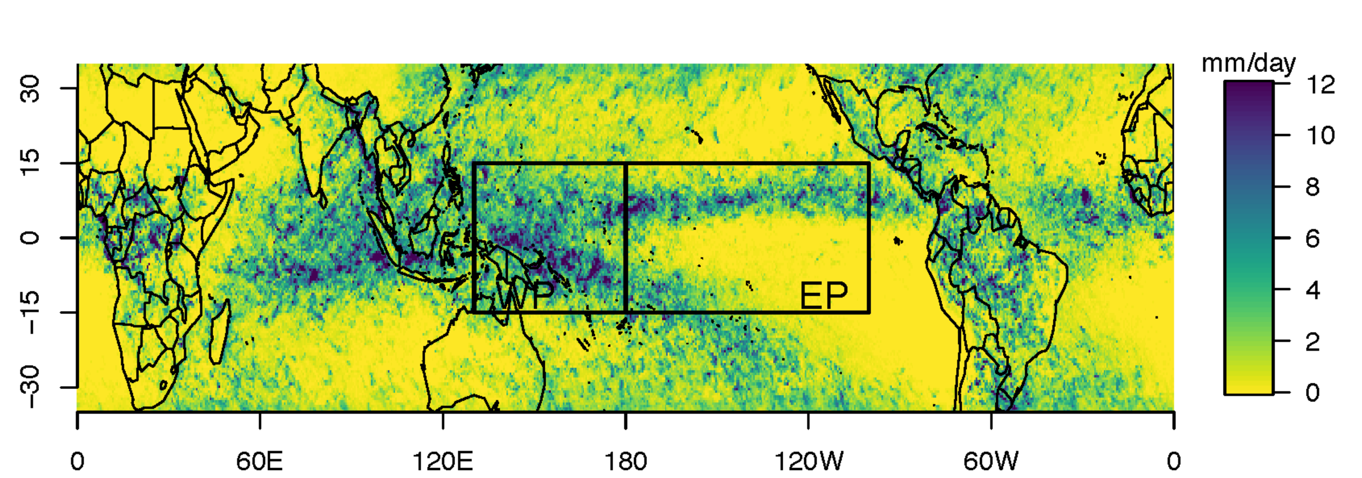


Figure 2. Satellite-measured annual average rainfall

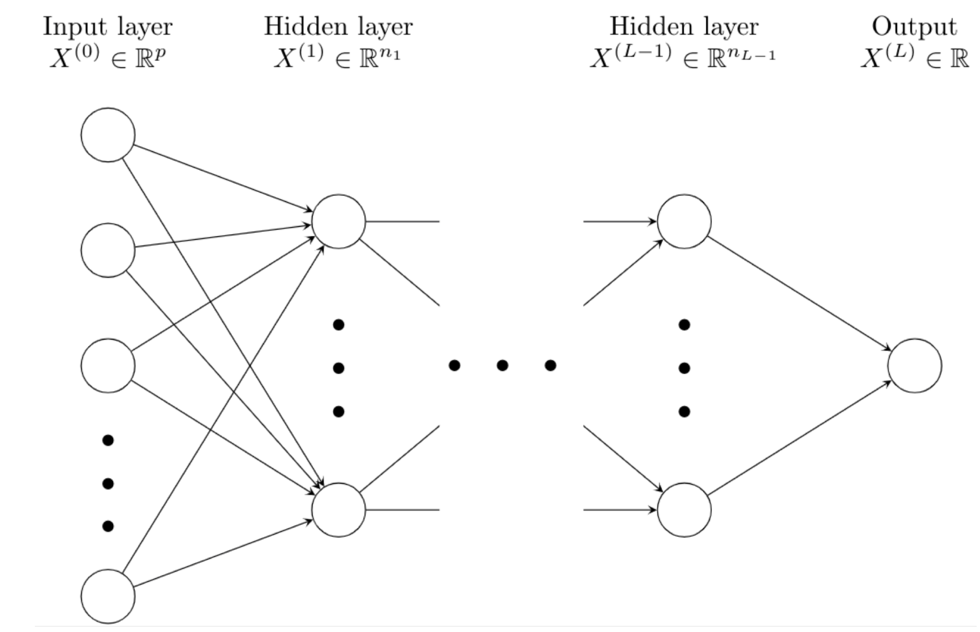


Figure 3. An example of a feedforward neural network

	Deep convective			Stratiform			Shallow convective		
	GLM	RF	NN	GLM	RF	NN	GLM	RF	NN
True Negative	**0.485**	**0.568**	**0.550**	**0.474**	**0.529**	**0.512**	**0.325**	**0.415**	**0.361**
False Negative	0.036	0.054	0.063	0.052	0.069	0.080	0.084	0.137	0.124
True Positive	**0.122**	**0.103**	**0.095**	**0.188**	**0.171**	**0.160**	**0.267**	**0.214**	**0.226**
False Positive	0.357	0.275	0.292	0.286	0.231	0.248	0.324	0.234	0.289
Total	1	1	1	1	1	1	1	1	1

Table 1. Prediction (classification) performance for each rain type (WP region)

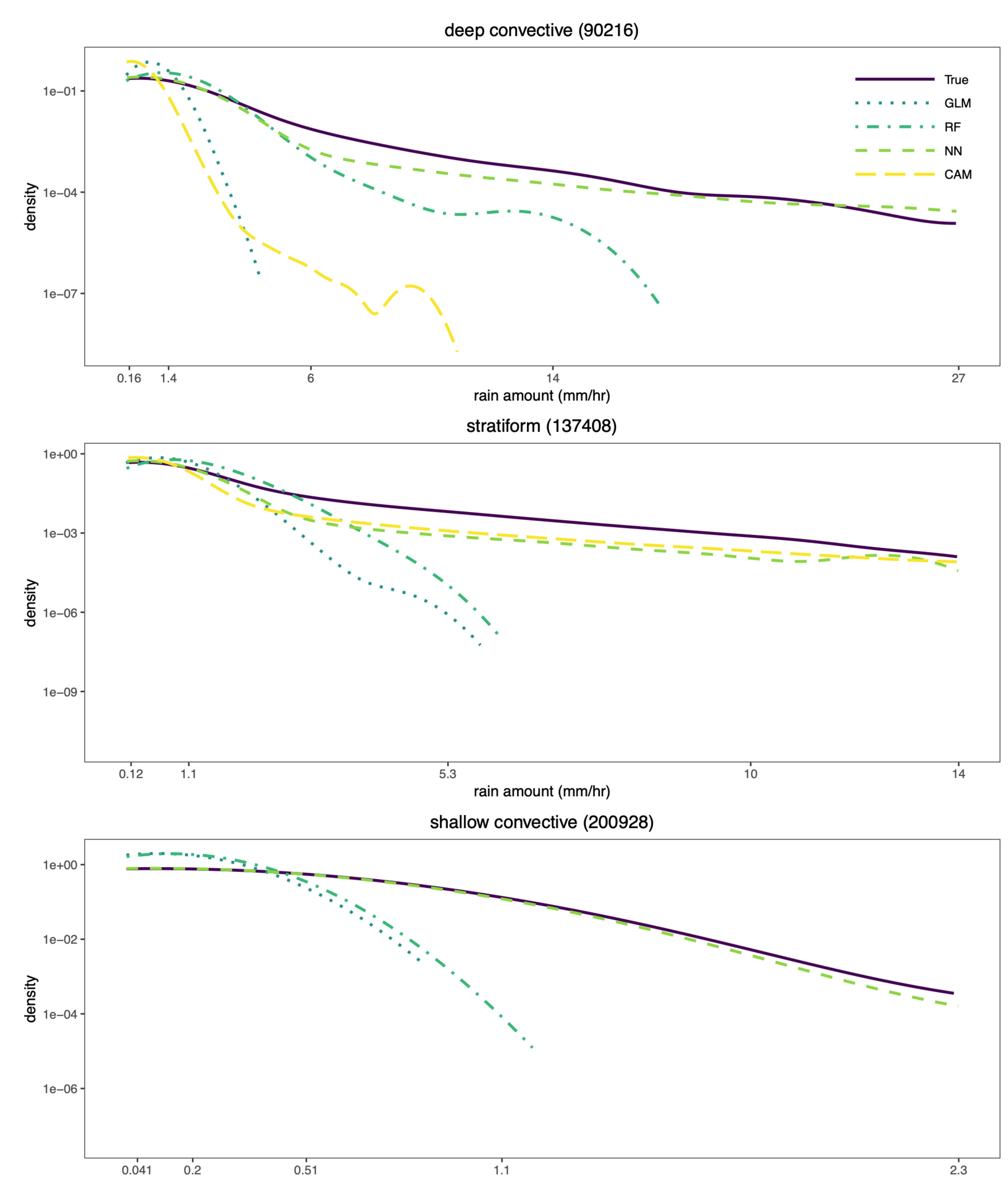


Figure 4. GPM observed and model predicted rain rate distributions: probability density vs. rain amount (WP region)

Models

- **Generalized linear model (GLM)**
  - Logistic regression for rain occurrence
  - Gamma regression for rain rate
- **Random Forest (RF)**
  - Based on decision trees
- **Neural Network (NN)**
  - 5-layer feed forward deep neural network (Figure 3)
- **Community Atmospheric Model (CAM)**

Conclusions

- All three methods performed well in predicting the occurrence of each of the three tropical building block rain types (Table 1).
- Due to the high complexity of the model structure, NN shows its advantage in characterizing the rain rate probability distributions well, even with the highly varying range of rain rates (Figure 4), performing much better than CAM
- However, high complexity raises the overfitting issue and can lead to "wrong" predictions. Compared to GLM, NN and RF are more flexible in modeling the response through a complicated function of all the predictors. But they are not as easy as GLM to interpret the results.

Acknowledgments

Texas A&M University T3 grant and Big Data Seed Grant

References

Yang, J., M. Jun, C. Schumacher, R. Saravanan, 2019: Predictive statistical representations of observed and simulated rainfall using generalized linear models. *Journal of Climate*, v.32, 3409-3427pp.

Wang, J., R. K. W. Wong, M. Jun, C. Schumacher, R. Saravanan, 2020: Statistical and machine Learning methods applied to the prediction of tropical rainfall. *Geophysical Research Letters*, submitted.